Tandemly duplicated *MYB* genes specifically in the Phaseoleae lineage are functionally diverged in the regulation of anthocyanin biosynthesis

Ruirui Ma[1], Wenxuan Huang[1], Quan Hu[1], Guo Tian[1], Jie An[1], Ting Fang[1], Jia Liu[1], Jingjing Hou[1], Meixia Zhao[2*], Lianjun Sun[1*]

[1] College of Agronomy and Biotechnology, China Agricultural University, Beijing 100193, China

[2] Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA

[*] Correspondences: Lianjun Sun (sunlj@cau.edu.cn); Meixia Zhao (meixiazhao@ufl.edu)

Abstract

Gene duplications have long been recognized as a driving force in the evolution of genes, giving rise to novel functions. The soybean genome is characterized by a large extent of duplicated genes. However, the extent and mechanisms of functional divergence among these duplicated genes in soybean remain poorly understood. In this study, we revealed that tandem duplication of *MYB* genes, which occurred specifically in the Phaseoleae lineage, exhibited a stronger purifying selection in soybean compared to common bean. To gain insights into the diverse functions of these *MYB* genes in anthocyanin biosynthesis, we examined the expression, transcriptional activity, metabolite, and evolutionary history of four *MYB* genes (*GmMYBA5*, *GmMYBA2*, *GmMYBA1* and *Glyma.09g235000*), which were presumably generated by tandem duplication in soybean. Our data revealed that *Glyma.09g235000* had become a pseudogene, while the remaining three *MYB* genes exhibited strong transcriptional activation activity and promoted anthocyanin biosynthesis in different soybean tissues. Furthermore, *GmMYBA5* produced distinct compounds in *Nicotiana benthamiana* leaves compared to *GmMYBA2* and *GmMYBA1* due to variations in their DNA binding domains. The lower expression of anthocyanin related genes in *GmMYBA5* resulted in lower levels of anthocyanins compared to

GmMYBA2 and *GmMYBA1*. Metabolomics analysis further demonstrated the diverse and differential downstream metabolites, suggesting their functional divergence in metabolites following gene duplication. Together, our data provided evidence of functional divergence within the *MYB* gene cluster following tandem duplication, which shed light on the potential evolutionary direction of gene duplications during legume evolution.

Keywords: tandemly duplicated *MYB* genes, anthocyanins, functional divergence, metabolites, legume

Introduction

In flowering plants, polyploidy or whole-genome duplication is a key process that leads to gene duplication, which provide genetic resources for generating functional novelty (Soltis et al., 2015; Soltis and Soltis, 2016). In addition to polyploidy, another two important processes that result in an increase in gene copy number are segmental and tandem duplication (Rizzon et al., 2006; Freeling, 2009). Following duplication events, duplicated genes undergo functional divergence, which can occur through pseudogenization resulting in loss of function, neofunctionalization leading to the acquisition of novel functions, or subfunctionalization where the duplicated genes retain partial functions of the ancestor gene (Sandve et al., 2018). The phenomenon of functional divergence in duplicated genes has been observed in many species. For instance, in citrus, the *Ruby2–Ruby1* gene cluster exhibits subfunctionalization, with these two genes exerting opposite effects in the regulation of anthocyanin biosynthesis (Huang et al., 2018). In maize, the duplicated *MYB* genes, *P1* and *P2*, show different expression patterns that contribute to tissue-specific pigmentation (Zhang et al., 2000). In *Solanum commersonii*, the tandem paralogs *ScAN1* and *ScAN2* have diverged in function with one gene specialized in anthocyanin production and the other one maintaining the conserved function of responding to cold stress (D'Amelia et al., 2018). These examples clearly demonstrate the functional divergence of genes following duplication events. Soybean, which has undergone two polyploidy events, has experienced substantial tandem duplication events, resulting in a

significant expansion of genes involved in the anthocyanin biosynthetic pathway (Kim et al., 2012).

Anthocyanins, one of the largest groups of plant flavonoid compounds, not only confer appealing colors to plants but also contribute their tolerance to biotic and abiotic stresses, including drought, cold, ultraviolet (UV)-B, heavy metals, herbivores, and pathogens (Gould, 2004; Hichri et al., 2011; Kovinich et al., 2015). Moreover, anthocyanins offer human health benefits to by protecting against chronic diseases such as metabolic syndrome, cardiovascular disease, and certain cancers (Zhang *et al.*, 2014; Putta et al., 2017). The biosynthesis of anthocyanins stems from the general phenylpropanoid pathway and involves several catalytic enzymes, including chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonoid 3'hydroxylase (F3'H), flavonoid 3'5' hydroxylase (F3'5'H), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), and UDP-flavonoid glucosyltransferase (UFGT) (Sundaramoorthy et al., 2015; Xu et al., 2015) (Supplementary Fig. S1). The expression of anthocyanin biosynthetic genes is primarily regulated by the MYB-bHLH (helix-loop-helix)-WD repeat (MBW) transcriptional complex (Albert et al., 2014; Lloyd et al., 2017; Wang et al., 2021; Wang et al., 2022). The critical part of the MBW transcriptional complex is MYB transcription factors, which are the key determinants of pigmentation and have been identified in various crops such as maize, rice, wheat and foxtail millet (Grotewold et al., 1994; Chin et al., 2016; Shin et al., 2016; Li et al., 2022b). Specific patterns and spatial localizations of anthocyanins determined by MYB transcription factors have also been reported in apple, snapdragon, petunia and lily (Liu et al., 2015). However, limited research has focused on the diverse metabolites and anthocyanin components, which play distinct roles in environmental adaptation produced by tandemly duplicated *MYB* genes.

Soybean (*Glycine max* (L.) Merr.) is one of the world's most important legume crops that provides plant protein, oil and other essential ingredients for humans and livestock (Malle *et al.*, 2020). Compared to its wild relative *Glycine soja* (Sieb. & Zucc.), cultivated soybean and landraces exhibit a wide range of morphological types

in order to meet human needs (Li et al., 2008). Various parts of the soybean plant, including hypocotyls, petioles, flowers, and seeds display significant natural variation in color due to distinct accumulation and distribution patterns of anthocyanins and proanthocyanins (Jeong et al., 2019; Xie et al., 2019). Previous studies have shown that the loss of anthocyanin pigment in cereals during domestication, diversification, and improvement is controlled by the same *MYB* gene (Li et al., 2022b). However, only a limited number of transcription factors responsible for anthocyanins biosynthesis have been studied in legume (Lu et al., 2021). The function and evolutionary history of syntenic block in legume is unknown. Therefore, it is crucial to investigate the mechanisms underlying the molecular mechanisms and functional divergence of tandemly duplicated MYB transcription factors in the regulation of anthocyanin biosynthesis in legumes.

Previous studies have described the collinear region of the MYB transcription factors cluster in legumes, which conservatively contributed to seed coat color in soybean, common bean and cowpea (Zabala et al., 2014; García-Fernández et al., 2021; Herniter et al., 2018). In this study, we revealed the tandemly duplicated *MYB* gene cluster that emerged prior to the divergence of soybean from other legume species during the common legume teraploidy event. We observed a stronger purifying selection acting on this cluster in soybean. With the exception of one pseudogene (*Glyma.09g235000*), all duplicated genes exhibited the potential to activate anthocyanin biosynthesis with tissue-specific patterns. Through ectopic expression and metabolomics analysis, we found that the *MYB* genes exhibited functional divergence in activation activities and specificity of downstream target catalytic enzymes, and metabolites. Additionally, a domain swap experiment suggested the divergent R2R3 DNA binding domain primary accounted for the distinct phenotype observed in *GmMYBA5*. Metabolomics analysis further demonstrated the diverse and differential downstream metabolites, suggesting their functional divergence in metabolites following gene duplication. Taken together, our findings shed light on the characteristics of functional divergence within a *MYB* gene cluster, and provide new insights into the evolutionary history of gene duplications

and functional divergence as a mechanism driving adaptation during legume evolution.

Results

Tandem duplication of the *MYB* genes occurs specifically in the Phaseoleae lineage and exhibits a stronger purifying selection in soybean

Soybean has undergone two polyploidy events and substantial tandem duplication events, resulting in a significant expansion of duplicated regions and genes involved in the regulation of anthocyanin biosynthesis, specifically the MYB transcription factors (Kim et al., 2012). The *R* locus (*GmMYBA2*, *Glyma.09g235100*), is an R2R3 MYB activator located on chromosome 9. It is flanked by three tandemly duplicated *MYB* genes: *GmMYBA5* (*Glyma.09g234900*), *GmMYBA1* (*Glyma.09g235300*) and *Glyma.09g235000* (Gao et al., 2021) (Fig. 1A). *Glyma.09g235000* was identified as a pseudogene due to its lack of expression in any soybean tissue (Gillman et al., 2011). To shed light on the evolutionary forces driving the divergence of the *MYB* genes, we analyzed a region containing the *MYB* gene cluster in different legumes genomes, which have been known to contributed conservatively to seed coat color in soybean, common bean, and cowpea (Zabala et al., 2014; García-Fernández et al., 2021; Herniter et al., 2018). The results indicated that the numbers of tandemly duplicated *MYB* genes varied across different species (Fig. 1B). In certain species such as *Trifolium pratense*, *Cicer arietinum*, *Pisum sativum* and *Aeschynomene evenia*, which diverged earlier from the legume common tetraploidy event, there was only a single copy of the *MYB* transcription factor gene in the collinear regions (Wang et al., 2017). However, an increased number of *MYB* genes was observed in the genera of *phaseolus* and *vigna*, which diverged from soybean approximately 27.3 million years ago (MYA). Notably, we detected four homologous genes in the syntenic region of common bean (*Phaseolus vulgaris*) (Fig. 1B). Interestingly, although all four homologous genes were present, the best matches for *GmMYBA5*, *GmMYBA2* and *GmMYBA1* in common bean were found to be *Phvul.008G038200*. To further explore the evolutionary divergence of these three soybean MYB genes, we next compared

the evolutionary distance of the three soybean *MYB* genes using common bean as an outgroup. Our data revealed that ω (Ka/Ks, where Ka represents nonsynonymous substitution and Ks represents synonymous substitution) for *GmMYBA5* was higher than that of *GmMYBA2* and *GmMYBA1* (Table 1), suggesting that *GmMYBA5* has experienced a lower intensity of purifying selection compared to the other two genes. Furthermore, our data indicated that the *MYB* genes in soybean exhibited lower rates of substitution for both Ka and Ks when compared to common bean, consistent with the genome-wide analysis of all genes between soybean and common bean (Zhao et al., 2017). Intriguingly, despite the significantly higher ω observed at the genome-wide level in soybean (Zhao et al., 2017), a lower ω was observed specially for these *MYB* genes in soybean, suggesting that the *MYB* genes have undergone a stronger purifying selection in soybean after their split with common bean.

It is worth noting that our data revealed that the soybean specific whole-genome duplication (13 MYA) did not significantly increase the number of tandemly duplicated genes in this region. Instead, it led to an increase in the number of homoeologous genes within the duplicated region on chromosome 18 (Fig. 1B). Interestingly, while the homoeologous genes of *GmMYBA2* and *GmMYBA1* on chromosome 18 were retained, the homoeologous gene of *GmMYBA5* was lost, likely occurring subsequent to the whole-genome duplication event. Taken together, our data suggest that the *MYB* gene cluster originated through tandem duplication prior to the soybean specific whole-genome duplication and *GmMYBA5* has undergone greater functional divergence. Anthocyanins, as important metabolites, are often induced by environmental stresses in plants (Li et al., 2022b). The increased number of *MYB* genes in different legumes may represent an adaptive strategy for these plants to respond to diverse external environmental conditions.

Tandemly duplicated *MYB* genes are putative anthocyanin synthesis regulators

The identification of *Glyma.09g235000* as a pseudogene (Gillman et al., 2011) prompted us to further investigate its function. To further confirm its role, we performed the ectopic expression of the coding sequence of *Glyma.09g235000* driven by the 35S promoter of cauliflower mosaic virus (CaMV) in *Arabidopsis thaliana*.

However, no phenotypic differences were observed in pigmentation between the transgenic plants and the wild types, indicating the loss of its regulatory function in anthocyanin synthesis (Supplementary Fig. S2). Therefore, we focused on the molecular function of the remaining three genes. Alignment of the amino acid sequences showed an 86.58% similarity among *GmMYBA5*, *GmMYBA2* and *GmMYBA1*. These three genes all contained the conserved R2 and R3 repeats, as well as the [D/E]Lx2[R/ K]x3Lx6Lx3R domain, which interacts with the R/B-like bHLH proteins (Supplementary Fig. S3). In addition, the conserved motif KPRPR[S/T] [F/L], which is important for anthocyanin activation, were present in all three genes (Stracke et al., 2001) (Fig. 2A). To elucidate the evolutionary relationship of these *MYB* genes, a phylogenetic tree was constructed using the protein sequences of known anthocyanin and proanthocyanin related MYB regulators from *Arabidopsis*, *Medicago* and *Vinifera*. The phylogenetic analysis classified these homologous genes into two major groups "anthocyanin activator" and "proanthocyanin activator". Notably, all of the three *MYB* genes from the cluster fell within the "anthocyanin activator" group, indicating their potential roles in anthocyanin biosynthesis (Fig. 2B).

GmMYBA5, _GmMYBA2_ and _GmMYBA1_ are transcription factors that can activate anthocyanin biosynthesis

The subcellular distribution of GmMYBA5, GmMYBA2 and GmMYBA1 was investigated by expressing fusion protein of these genes with green fluorescent protein (GFP) in tobacco (*Nicotiana benthamiana*) leaf epidermal cells. The GFP fluorescent signals were observed specifically in the nuclei of all of the three genes (Fig. 3A), indicating that these genes are localized in the nucleus. To access their transcriptional activity, the full coding sequences of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* were fused with the GAL4 DNA binding domain of yeast in the expression vector pGBKT7. The results showed that yeast transformants carrying these constructs were able to grow on medium lacking Trp, His and Ade, indicating that these three genes function as transcription factors with robust transcriptional activity (Fig. 3B). Furthermore, we introduced overexpression constructs containing the coding sequences of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* driven by the 35S

promoter of cauliflower mosaic virus (CaMV) into *Arabidopsis thaliana*. Compared to the wild type control, all five independent T_2 transgenic lines exhibited purple pigment accumulation in various tissues, including seedlings, leaves, roots, and leaf veins (Fig. 3C and Supplementary Fig. S4). These findings demonstrate that *GmMYBA5*, *GmMYBA2* and *GmMYBA1* are nucleus-localized transcription factors with strong transcriptional activity that can activate the anthocyanin biosynthesis in vivo.

Functional divergence of *GmMYBA5*, *GmMYBA2* and *GmMYBA1*

Previous studies have reported the specific expression pattern of *GmMYBA2* in the seed coat (Gillman et al., 2011). To validate this, quantitative reverse transcriptase (RT)-PCR analysis was performed on various soybean tissues, including stems, leaves, pods, and seeds. Consistent with previous findings, *GmMYBA2* was specifically expressed in seeds. Conversely, *GmMYBA5* was primarily expressed in vegetative tissues such as stems, leaves, and pods. In contrast, *GmMYBA1* displayed low expression levels restricted to leaves and stems (Fig. 4A). To further investigate the impact of ectopic expression of these three genes, transient overexpression of these three genes were conducted in *Nicotiana benthamiana* leaves. Leaves overexpressing *GmMYBA5*, *GmMYBA2* and *GmMYBA1* exhibited a noticeable anthocyanin-pigmented phenotype, and the anthocyanin content was significantly higher compared to the leaves with an empty vector control (Fig. 4B and C). Interestingly, the anthocyanin solution extracted from leaves overexpressing *GmMYBA5* displayed a brownish color instead of the purple color observed in leaves overexpressing *GmMYBA2* and *GmMYB1* (Fig. 4B). This observation suggests that the ectopic expression of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* leads to the production of distinct metabolites in *Nicotiana benthamiana* leaves.

To elucidate the molecular basis underlying these different metabolites, domain swap experiments were performed. The DNA binding domains (BD; corresponding to R2R3 domain) and the activating domains (AD; corresponding to C-terminal domain) of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* were exchanged, generating chimeric proteins (Fig. 5A). Our results showed that when the BD of *GmMYBA5* was fused

with the AD of either *GmMYBA2* or *GmMYBA1*, the resulting chimeric proteins retained the phenotype of *GmMYBA5* (Fig. 5B). This indicates that the unique brownish metabolites produced by *GmMYBA5* is primarily influenced by its divergent R2R3 domains and potentially by its binding ability.

Additional experiments were performed in soybean hair roots to further investigate the function of the MYB genes. The highly expressed GmScreamM4 promoter (pM4) was utilized to replace 35S promoter of the plant expression vector pTF101, (Zhang et al., 2015), and the full coding sequences of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* were inserted. These constructs were transformed into *Agrobacterium rhizogenes* K599, and soybean cotyledons were infected to generate transgenic soybean hairy roots. Transgenic soybean hairy roots carrying the three genes exhibited purple pigmentation in some roots, with the pM4:*GmMYBA5* transgenic soybean hairy roots displaying less pigmentation compared to the others (Fig. 5C). Expression analysis revealed that the expression levels of catalytic enzymes genes, including *GmF3H*, *GmF3'H*, *GmF3'5'H*, *GmDFR1*, *GmDFR2*, *GmANS1*, *GmANS2*, *GmUGT78K1*, and *GmUGT78K2*, were significantly upregulated in the overexpression lines, although to a lesser extent in *GmMYBA5* lines compared to the other lines (Fig. 5D). These results suggest that *GmMYBA5* has a lower capacity to activate catalytic genes involved in anthocyanin biosynthesis, resulting in reduced visible anthocyanin accumulation in roots. Overall, our data demonstrate that these tandemly duplicated *MYB* genes have undergone functional divergence, leading to tissue-specific expression patterns and differential activation abilities for anthocyanin biosynthesis.

Metabolomics analysis of metabolites produced by *GmMYBA5*, *GmMYBA2* and *GmMYBA1*

To gain further insights into the functionality of the three *MYB* genes, we examined the downstream compounds in soybean hairy roots overexpressing pM4:*GmMYBA5,* pM4:*GmMYBA2* and pM4:*GmMYBA1*, as well as control roots with an empty vector, using Liquid Chromatograph Mass Spectrometer (LC-MS). In total, we identified 156 flavonoid compounds across all samples, with 80 (51.3%) belonging to flavones and flavonols, 30 (19.2%) classified as isoflavonoids, and 12 (7.7%) identified as

anthocyanins (Fig. 6A and Supplementary Table S1). To ensure the reliability of the metabolite extraction and detection, we assessed the total ion flow diagram (TIC) of the mass spectrometry analysis and calculated the Pearson correlation coefficient for different quality control (QC) samples. Our data demonstrated a high correlation, indicating the repeatability of both metabolite extraction and detection (Supplementary Fig. S5A and B). Furthermore, we performed principal component analysis (PCA) and hierarchical clustering to evaluate the biological repeatability among the samples. The PCA score plot and hierarchical clustering heatmap clearly revealed that all samples were distinctly separated into the four expected groups, affirming the high quality and repeatability of the data (Supplementary Figs. S5C and S6).

The abundances of the 156 metabolic compounds were compared between the overexpression groups of the three genes and the control group with empty vectors, as well as within the overexpression groups. Differential metabolites were defined as those meeting the threshold criteria of $p \leq 0.05$ (Student's *t* test) and $|\log_2 FC$ (Fold Change) $| \geq 1$. Our analysis revealed that the overexpression of *GmMYBA2* and *GmMYBA1* identified 72 and 81 differential metabolites, respectively, whereas *GmMYBA5* overexpression resulted in 37 differential metabolites (Fig. 6B). Among these differential metabolites, 22, 52, and 64 were upregulated, while 15, 20 and 17 were down-regulated when comparing the overexpression of *GmMYBA5*, *GmMYBA2* and *GmMYBA1*, respectively, with the empty vector (Fig. 6B, Supplementary Fig. S7A-F). Notably, the numbers of up-regulated flavonoids metabolites were greater than the number of down-regulated metabolites. Specially, out of the 37 differential metabolites driven by the overexpression of *GmMYBA5*, 22 (59.5%) were up-regulated, which is significantly lower than the numbers of the upregulated metabolites driven by the overexpression of *GmMYBA2* (72.2%) and *GmMYBA1* (79.0%) (Fig 6B). Among these up- and down-regulated metabolites, 20 (18.9%) were shared by all of the three genes (Fig. 6C and Supplementary Table S2). Furthermore, of the 72 differential metabolites driven by the overexpression of *GmMYBA2*, 61 (84.7%) were shared with *GmMYBA1*, suggesting potential redundancy in the

298 regulation of flavonoid synthesis between *GmMYBA2* and *GmMYBA1* (Fig. 6B and
299 C).

300 To further explore the regulatory role of these three *MYB* genes in anthocyanin
301 biosynthesis, we focused on the analysis of 12 anthocyanins. These 12 anthocyanins
302 encompassed cyanidin derivatives (idaein chloride, cyanin chloride, cyanidin
303 O-rutinoside, cyanidin 3-O-glucoside, keracyanin chloride), delphinidin derivatives
304 (myrtillin chloride, delphinidin chloride), pelargonidin derivatives (pelargonidin
305 chloride, callistephin chloride), and petunidin derivatives (petunidin 3-O-rutinoside,
306 petunidin 3-O-glucoside) (Fig. 6D). Our data showed that the majority of these 12
307 anthocyanins were upregulated in the overexpression lines of the three *MYB* genes.
308 Specially, in the *GmMYBA1* overexpression line, 10 (83.3%) out of the 12
309 anthocyanins were upregulated (Fig. 6D). In contrast, *GmMYBA5* overexpression
310 resulted in fewer and lower levels of anthocyanins in soybean hairy roots, which
311 could explain the reduced pigmentation observed in these roots. This observation is
312 consistent with the expression levels of catalytic enzymes genes (Fig. 6D). It is worth
313 noting that glycoside derivatives constituted a significant proportion (40.91% and
314 44.19%) of the differential metabolites between *GmMYBA5* and *GmMYBA2*, as well
315 as between *GmMYBA5* and *GmMYBA1* (Fig. 6D and Supplementary Table S3). These
316 data suggest that *GmMYBA5* has partially lost its function in regulating the expression
317 of genes involved in anthocyanin biosynthesis, resulting in reduced production of
318 glycoside derivatives. On the other hand, *GmMYBA2* and *GmMYBA1* appear to have
319 redundant functions in regulating the synthesis of flavonoids compounds including
320 anthocyanins.

321 ## Discussion

322 ### MYB transcription factors are possible hotspots for tandem duplication

323 Tandem duplication represents one of the key processes by which the copy number of
324 genes can be increased, leading to the emergence of new genetic resources during the
325 course of evolutionary history in many organisms. Certain classes, such as *R* genes,
326 are known to be prone to tandem duplication (Meyers et al., 2003; Innes et al., 2008).

It appears that *MYB* genes are also hotspots for tandem duplication in soybean, as well as in several other species (O'Neil et al., 2007; Wang et al., 2023). Previous studies have demonstrated that selection is relaxed on tandemly duplicated genes relative to non-tandemly duplicated genes. This relaxed selection may allow more transposon insertions or less efficient purging of transposons near these genes (Zhao et al., 2017). However, this does not seem to be the case for the three *MYB* genes examined in our study. Although these *MYB* genes have undergone functional divergence, they still retain the function of the ancestral gene in regulating anthocyanin biosynthesis. According to the gene dosage balance hypothesis, genes encoding proteins that interact with other proteins are more sensitive to changes in order to maintain the overall network functionality (Birchler and Veitia, 2021). Given that these *MYB* genes function as transcription factors, they likely interact with other proteins, which could explain their conserved function. Nevertheless, our study reveals a tissue-specific pattern for these gene (Fig. 4A), a characteristic commonly observed in many other tandemly duplicated genes (Zhao et al., 2017).

To understabnd the mechanism underlying MYB tandem duplication, we also examined the transposon sequences within the genic and flanking regions of the three *MYB* genes. Surprisingly, we found very few transposon sequences enriched in these regions. However, considering the high rate of transposon turnover in soybean, it is likely that the elements that contribute to the tandem duplication of the *MYB* genes have already been eliminated from the genome. Further investigation could involve examining other recently tandemly duplicated *MYB* gene clusters to assess the role of transposons in mediating the tandem duplication of *MYB* genes.

Functional divergence of the *MYB* duplicated genes

Subfunctionalization proposes that duplicated genes originating from a common ancestor specialize in complementary functions to maintain the original function of their ancestral gene (Sandve et al., 2018). In our research, the tandemly duplicated *MYB* genes have undergone both pseudogenization and subfunctionalization. A putative pseudogene (*Glyma.09g235000*) structurally resembled the gene but lacked expression and function. Other genes within the *MYB* gene cluster acted as activators

of anthocyanin synthesis but exhibited specialization in tissue-specific expression patterns, varied activation abilities, and diverse downstream metabolites (Fig. 7). Conserved MYB transcription factors contain activation domain in the C-terminal and A binding domain in the N-terminal, enabling them to regulate anthocyanins synthesis by forming MYB-bHLH-WD ternary complexes. (Wang et al., 2021). We have demonstrated that variations in the BD contribute to the divergent anthocyanin productions. The conserved bHLH-interacting motif [D/E]Lx2[R/ K]x3Lx6Lx3R is present in GmMYBA5, GmMYBA2 and GmMYBA1, albeit with two substitutions of L (Leucine) to M (Methionine) (Liu et al., 2015) (Supplementary Fig. S3). However, whether these three genes exhibit distinct interaction preferences for basic helix-loop-helix proteins, thereby resulting in the different metabolites, remains to be determined.

GmMYBA2 plays a regulatory role in the late stage of anthocyanin biosynthesis by controlling enzymes such as flavonoid 3'5'hydroxylase, dihydroflavonol 4-reductase, anthocyanidin synthase and glucosyltransferase, which are involved (Gao et al., 2021). In contrast, overexpression of *GmMYBA5* led to reduced anthocyanin production, indicating a partial loss of its regulatory function in regulating catalytic enzymes genes. On the other hand, *GmMYBA2* and *GmMYBA1* exhibit functional redundancy in regulating anthocyanidin synthesis, although they display distinct expression patterns (Fig. 3A). Furthermore, the expression of *GmMYBA5* and *GmMYBA1* is generally induced by abiotic and biotic stresses, as inferred from publicly available RNA-seq libraries (http://ipf.sustech.edu.cn/pub/soybean/), suggesting their putative function in adapting to changing environment factors (Zhang et al., 2020) (Supplementary Fig. S8A-C). These results offer additional insights into the evolutionary trajectory and function divergence resulting from gene duplication in the regulation of plant secondary metabolites.

Insights into the relationship of gene duplication and environmental adaptation

Gene duplication is an important force for genome evolution and environmental adaptation (Magadum et al., 2013; Chen et al., 2022b). Numerous studies have highlighted the occurrence of adaptive gene duplications in response to environmental

stresses (Brown et al., 1998; Riehle et al., 2001; Hastings, 2007; James et al., 2008). In our investigation of the collinear regions of the MYB transcription factors cluster in legumes, we found that these orthologous *MYB* genes contribute to seed coat color in common bean and cowpea (García-Fernández et al., 2021; Herniter et al., 2018). The conservation of MYB transcription factors across species suggests their essential role in enabling efficient responses to environmental stresses (Saigo et al., 2020). The observed copy-number variation and functional divergence of *MYB* genes in legumes serve as a compelling example of adaptive gene duplications in the evolutionary history of legumes, which enhances our understanding of the role of gene duplications as a mechanism of adaptation.

Anthocyanin can be used as a morphological marker

As Anthocyanin pigmentation can be identified by naked eyes, it has been employed as a visible marker in the maize haploid inducer system, facilitating the efficient selection of haploid embryos (Chen et al., 2022a). Selecting transgenic tissues and plants during plant genetic transformation is a laborious and time-consuming process. Recently, R2R3 type MYB transcription factors have been utilized as visible markers for the selection of transformed tissues and plants in various species (Zhang et al., 2019; Huang et al., 2021; Lim et al., 2022). Soybean hairy roots can be obtained by *Agrobacterium rhizogenes* mediated transformation, forming chimera plants with transformed hairy roots and untransformed shoots. In our study, we constructed plant expression vectors containing the GmScreamM4 promoter (pM4), which drove the expression of *GmMYBA2* and *GmMYBA1* in transgenic soybean hairy roots, leading to the manifestation of a purple color (Fig. 5C). The application of *GmMYBA2* and *GmMYBA1* as strong activators of anthocyanin biosynthesis in soybean holds promising prospects for their utilization as markers to facilitate the effective identification of transformed hairy roots.

Materials and Methods

Plant materials

The soybean cultivar Wm82 (Williams 82) was grown in the greenhouse at China

Agriculture University. Plant tissues were collected at the reproductive growth stage R4 (full pod) and immediately frozen into liquid nitrogen before being stored at -80 °C. The *Arabidopsis thaliana* ecotype Columbia was used for genetic transformation. The transient ectopic expression material, tobacco (*Nicotiana benthamiana*), was grown in the incubator with the constant temperature of 25℃.

Estimation of evolutionary distance and analysis of micro-synteny

To estimate evolutionary distance, we followed the method previously described (Zhao et al. 2017; Yin et al. 2022). In brief, the coding sequences of the homologous genes were aligned using ClustalW (Thompson et al., 1994) with default parameters, followed by manual inspection. Pairwise alignments of the homologous genes were performed to calculate Ka and Ks using the yn00 module under the PAML software (Yang, 2007). Micro-synteny analysis was conducted using the LegumeInfo database (https://www.legumeinfo.org/). The divergence time was obtained from the TimeTree database (https://timetree.org/).

cDNA synthesis and quantitative real-time PCR analysis

Total RNA was isolated from soybean leaves, pods, stems and seeds by the StarSpin HiPure Plant RNA Mini Kit (GenStar). To eliminate any genomic DNA contamination, the StarScriptII First-strand cDNA Synthesis Mix With gDNA Remover (GenStar) was employed to synthesize the first-strand complementary DNA (cDNA), following the manufacturer's instructions. Quantitative real-time PCR (qRT-PCR) analysis was performed using 2×RealStar Green Fast Mixture (with ROX II) (GenStar) according to the manufacturer's protocol, and amplification was carried out using the ABI 7500 Real-time PCR system (Applied Biosystems, USA). The *Actin11* gene was used as the internal control. The qRT-PCR data were analysed using the $2^{-\Delta\Delta Ct}$ analysis method. Details of all the primers can be found in Supplementary Table S4.

Analyses of phylogenetic relationship

Sequence alignment of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* was performed by DNAMAN software version 8.0. A phylogenetic tree was constructed using the Neighbor-Joining method in the MEGA7 program. The statistical significance of individual nodes was assessed by bootstrap analysis with 1,000 replicates.

Arabidopsis thaliana **transformation**

The full-length coding sequences of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* cloned from Wm82 and the synthesized code sequence of *Glyma.09g235000* were introduced into the pTF101 vector. The coding sequences were driven by the 35S promoter of cauliflower mosaic virus (CaMV). The resulting constructs were then introduced into *Agrobacterium* strain GV3101 for transformation of *Arabidopsis thaliana* using the floral dip method (Bent, 2006). The presence of the constructs in the transgenic plants was confirmed by PCR, followed by the sequencing of the PCR fragment with specific primers. All primers used are listed in Supplementary Table S4.

Subcellular localization

The full-length CDS of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* were cloned from Wm82. These CDS sequences were fused in-frame with the GFP coding sequence and subsequently inserted into a vector through combinational joining. The resulting constructs were introduced into *Agrobacterium* strain GV3101 for infection of *Nicotiana benthamiana* leaves. The GFP fluorescent signals emitted by the tobacco leaf tissues were captured using a Zeiss confocal laser scanning microscope (Zeiss, Germany).

Transcriptional activation activity assay

The transactivation activity assay was performed as previously described (Hou et al., 2022). In brief, the coding sequences of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* were fused with the GAL4 DNA-binding domain (BD) in the plasmid pGBKT7. These constructs were subsequently transformed into the yeast strain AH109 according to the procedure previously described (Gietz et al., 2007). The yeast colonies were then patched onto SD/-Trp and SD/-Trp/-His/-Ade plates and incubated at 30°C for 3 days.

Transient ectopic expression in *Nicotiana benthamiana*

The pTF-gene constructs were used to perform transient ectopic expression in *Nicotiana benthamiana*, follwoing the procedure described earlier in the subcellular localization section. After a 5-day incubation period, the phenotype of the infiltrated leaf materials was observed, and the extracted anthocyanins were measured.

Soybean hairy root transformation

The 35S promoter of cauliflower mosaic virus (CaMV) in the plant expression vector pTF101 was substituted with the pM4 promoter of soybean. The plasmids were then transformed into *Agrobacterium rhizogenes* K599 for infection of soybean cotyledons to generate transgenic soybean hairy roots. The procedure for transformation and infection followed the method previously described with slight modifications (Kereszt et al., 2007; Guo et al., 2011).

Anthocyanin extractions and measurements

Total anthocyanin contents were determined as previously described (Huang et al., 2018). Plant tissues were collected and subsequently ground into powders using liquid nitrogen. A total of 0.2 g powders was extracted in 1 ml of methanol containing 0.1% (v/v) HCl, followed by incubation on ice for 30 min. After centrifugation at 12,000 rpm for 10 min, the supernatant was measured at 530 nm (A530) and 657 nm (A657) using a spectrophotometer (SOPTOP, China). The relative content of anthocyanin was calculated using the formula (A530-0.25×A657)/fresh weight.

Metabolomics analysis

Samples were prepared and metabolites were extracted following the protocols provided by Novogene Co., Ltd. (Beijing, China). The extraction solution was then injected into the LC-MS/MS system. LC-MS/MS analyses were performed using an ExionLC™ AD system (SCIEX) coupled with a QTRAP® 6500+ mass spectrometer (SCIEX) at Novogene Co., Ltd. (Beijing, China). The detection of the experimental samples using Multiple Reaction Monitoring (MRM) was based on the in-house database of Novogene. The generated data files from the HPLC-MS/MS were processed using the SCIEX OS Version 1.4 to integrate and correct the peaks. Sample normalization and significantly differential accumulated flavonoids were analyzed using MetaboAnalyst 5.0 (https://www.metaboanalyst.ca/). Samples were normalized by sum and the data were transformed using log10. The standard threshold criteria were set as follows: $p \leq 0.05$ (Student's *t* test) and $|\log_2 FC| \geq 1$. Principal component analysis (PCA) and hierarchical clustering were carried out using MetaboAnalyst 5.0 (https://www.metaboanalyst.ca/).

Funding

This work was supported by the National Natural Science Foundation of China (grant nos. 32072089 and 31871708), the Recruitment Program of Global Experts and the Fundamental Research Funds for the Central Universities (2022TC021).

Author Contributions

L.S. and M.Z. conceived and designed the project. R.M., Q.H., G.T. and J.L. performed the experiments and W.H., J.A., T.F., and J.H. analysed the data. R.M. wrote the manuscript. M.Z. and L.S. revised the paper. All authors read and approved of this manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest associated with this work.

References

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell. 2014:**26(3)**:962-980. https://doi.org/10.1105/tpc.113.122069

Bent A. Arabidopsis thaliana floral dip transformation method**.** Methods Mol Biol, 2006:**343**:87-103. https://doi.org/10.1385/1-59745-130-4:87

Birchler JA, Veitia RA. One hundred years of gene balance: how stoichiometric issues affect gene expression, genome evolution, and quantitative traits. Cytogenet Genome Res, 2021:**161**(10-11):529-550. https://doi.org/10.1159/000519592

Brown CJ, Todd KM, Rosenzweig RF. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol, 1998:**15**(8):931-942. https://doi.org/10.1093/oxfordjournals.molbev.a026009

Chen C, Liu XQ, Li SZ, Liu CX, Zhang YL, Luo LL, Miao LQ, Yang WZ, Xiao

ZJ, Zhong Y, et al. Co-expression of transcription factors *ZmC1* and *ZmR2* establishes an efficient and accurate haploid embryo identification system in maize. Plant J, 2022a:**111**(5):1296-1307. https://doi.org/10.1111/tpj.15888

Chen Y, Fang T, Su H, Duan SF, Ma RR, Wang P, Wu L, Sun WB, Hu QC, Zhao MX, et al. A reference-grade genome assembly for *Astragalus mongholicus* and insights into the biosynthesis and high accumulation of triterpenoids and flavonoids in its roots. Plant Commun, 2022b:**4**:100469. https://doi.org/10.1016/j.xplc.2022.100469

Chin HS, Wu YP, Hour AL, Hong CY, Lin YR. Genetic and evolutionary analysis of purple leaf sheath in rice. Rice, 2016:**9**:8. https://doi.org/10.1186/s12284-016-0080-y

D'Amelia V, Aversano R, Ruggiero A, Batelli G, Appelhagen I, Dinacci C, Hill L, Martin C, Carputo D. Subfunctionalization of duplicate MYB genes in *Solanum commersonii* generated the cold-induced *ScAN2* and the anthocyanin regulator *ScAN1*. Plant Cell and Environ, 2018: **41**(5):1038-1051. https://doi.org/10.1111/pce.12966

Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev of Plant Biol, 2009:**60**:433-453. https://doi.org/10.1146/annurev.arplant.043008.092122

Gao RF, Han TT, Xun HW, Zeng XS, Li PH, Li YQ, Wang YN, Shao Y, Cheng X, Feng XZ, et al. MYB transcription factors *GmMYBA2* and *GmMYBR* function in a feedback loop to control pigmentation of seed coat in soybean. J Exp Bot, 2021:**72**(12):4401-4418. https://doi.org/10.1093/jxb/erab152

García-Fernández C, Campa A, Ferreira JJ. Dissecting the genetic control of seed coat color in a RIL population of common bean (*Phaseolus vulgaris* L.). Theor and Appl Genet, 2021:**134**(11):3687-3698. https://doi.org/10.1007/s00122-021-03922-y

Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc, 2007:**2**(1):31-34. https://doi.org/10.1038/nprot.2007.13

Gillman JD, Tetlow A, Lee JD, Shannon JG, Bilyeu K. Loss-of-function mutations affecting a specific *Glycine max* R2R3 MYB transcription factor result in brown

564 hilum and brown seed coats. BMC Plant Biol, 2011:**11**:155 https://doi.org/155.
565 10.1186/1471-2229-11-155
566 **Gould KS.** Nature's Swiss Army Knife: The diverse protective roles of anthocyanins
567 in leaves. J of Biomed and Biotechnol, 2004:**2004**(5):314-320.
568 https://doi.org/10.1155/S1110724304406147
569 **Grotewold E, Drummond BJ, Bowen B, Peterson T.** The MYB-homologous
570 *p-gene* controls phlobaphene pigmentation in maize floral organs by directly
571 activating a flavonoid biosynthetic gene subset. Cell, 1994:**76**(3):543-553.
572 https://doi.org/ 10.1016/0092-8674(94)90117-1
573 **Guo WB, Zhao J, Li XX, Qin L, Yan XL, Liao H.** A soybean β-expansin gene
574 *GmEXPB2* intrinsically involved in root system architecture responses to abiotic
575 stresses. Plant J, 2011:**66**(3):541-552.
576 https://doi.org/10.1111/j.1365-313X.2011.04511.x
577 **Hastings PJ.** Adaptive amplification. Crit Rev Biochem and Mol Biol,
578 2007:**42**(4):271-283. https://doi.org/10.1080/10409230701507757
579 **Herniter IA, Muñoz-Amatriaín M, Lo S, Guo YN, Close TJ.** Identification of
580 candidate genes controlling black seed coat and pod tip color in cowpea (*Vigna*
581 *unguiculata* [L.] Walp). G3 (Bethesda), 2018:**8**(10):3347-3355.
582 https://doi.org/10.1534/g3.118.200521
583 **Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V.** Recent advances in
584 the transcriptional regulation of the flavonoid biosynthetic pathway. J Exp Bot,
585 2011:**62**(8):2465-2483. https://doi.org/10.1093/jxb/erq442
586 **Hou JJ, Fan WW, Ma RR, Li B, Yuan ZH, Huang WX, Wu YY, Hu Q, Lin CJ,**
587 **Zhao XQ, et al.** *MALE STERILITY 3* encodes a plant homeodomain-finger protein for
588 male fertility in soybean. J Integr Plant Biol, 2022:**64**(5):1076-1086.
589 https://doi.org/10.1111/jipb.13242
590 **Huang D, Wang X, Tang ZZ, Yuan Y, Xu YT, He JX, Jiang XL, Peng SA, Li L,**
591 **Butelli E, et al.** Subfunctionalization of the *Ruby2-Ruby1* gene cluster during the
592 domestication of citrus. Nat Plants, 2018:**4**(11):930-941.
593 https://doi.org/10.1038/s41477-018-0287-6

Huang T, Xin S, Fang Y, Chen T, Chang J, Ko N. Huang H, Hua Y. Use of a novel R2R3-MYB transcriptional activator of anthocyanin biosynthesis as visual selection marker for rubber tree (*Hevea brasiliensis*) transformation. Ind Crop and Prod, 2021:**174**:114225. https://doi.org/10.1016/j.indcrop.2021.114225

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NW, Couloux A, Dalwani A, Denny R, et al. Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiology, 2008:**148**(4):1740-1759. https://doi.org/10.1104/pp.108.127902

James TC, Usher J, Campbell S, Bond U. Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. Curr Genet, 2008:**53**(3):139-152. https://doi.org/10.1007/s00294-007-0172-8

Jeong SC, Moon JK, Park SK, Kim MS, Lee K, Lee SR, Jeong N, Choi MS, Kim N, Kang ST, *et al.* Genetic diversity patterns and domestication origin of soybean. Theor Appl Genet, 2019:**132**(4):1179-1193. https://doi.org/10.1007/s00122-018-3271-7

Kereszt A, Li D, Indrasumunar A, Nguyen CD, Nontachaiyapoom S, Kinkema M, Gresshoff PM. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. Nat Protoc, 2007:**2**(4):948-952. https://doi.org/10.1038/nprot.2007.141

Kim MY, Van K, Kang YJ, Kim KH, Lee SH. Tracing soybean domestication history: From nucleotide to genome. Breeding Sci, 2012:**61**(5):445-452. https://doi.org/10.1270/jsbbs.61.445

Kovinich N, Kayanja G, Chanoca A, Otegui MS, Grotewold E. Abiotic stresses induce different localizations of anthocyanins in *Arabidopsis*. Plant Signal Behav, 2015:**10**(7):1027850. https://doi.org/10.1080/15592324.2015.1027850

Li CX, Yu WJ, Xu JR, Lu XF, Liu YZ. Anthocyanin biosynthesis induced by MYB transcription factors in plants. Int J Mol Sci, 2022a:**23**(19):11701. https://doi.org/10.3390/ijms231911701

Li Y, Fang XJ, Lin ZW. Convergent loss of anthocyanin pigments is controlled by

the same MYB gene in cereals. J Exp Bot, 2022b:**73**(18):6089-6102. https://doi.org/10.1093/jxb/erac270

Li YH, Guan RX, Liu ZX, Ma YS, Wang LX, Li LH, Lin FY, Luan WJ, Chen PY, Yan Z, *et al.* Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. Theor Appl Genet, 2008:**117**(6):857-871. https://doi.org/10.1007/s00122-008-0825-0

Lim SH, Kim DH, Cho MC, Lee JY. Chili Pepper *AN2 (CaAN2)*: A visible selection marker for nondestructive monitoring of transgenic plants. Plants (Basel), 2022.**11**(6):820. https://doi.org/10.3390/plants11060820

Liu JY, Osbourn A, Ma PD. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. Mol Plant, 2015:**8**(5):689-708. https://doi.org/10.1016/j.molp.2015.03.012

Lloyd A, Brockman A, Aguirre L, Campbell A, Bean A, Cantero A, Gonzalez A. Advances in the MYB-bHLH-WD Repeat (MBW) pigment regulatory model: Addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. Plant and Cell Physiol, 2017: **58**(9):1431-1441. https://doi.org/10.1093/pcp/pcx075

Lu N, Rao X, Li Y, Jun JH, Dixon RA. Dissecting the transcriptional regulation of proanthocyanidin and anthocyanin biosynthesis in soybean (*Glycine max*). Plant Biotechnol J, 2021:**19**(7):1429-1442. https://doi.org/10.1111/pbi.13562

Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. J Genet, 2013:**92**(1):155-161. https://doi.org/10.1007/s12041-013-0212-8

Malle S, Morrison M, Belzile F. Identification of loci controlling mineral element concentration in soybean seeds. BMC Plant Biol, 2020:**20**(1):419. https://doi.org/10.1186/s12870-020-02631-w

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell, 2003:**15**(4):809-834. https://doi.org/10.1105/tpc.009308

O'Neil J, Tchinda J, Gutierrez A, Moreau L, Maser RS, Wong KK, Li W,

McKenna K, Liu XS, Feng B, et al. Alu elements mediate MYB gene tandem duplication in human T-ALL. J Exp Med, 2007:**204**(13):3059-3066. https://doi.org/10.1084/jem.20071637

Putta S, Yarla NS, Peluso I, Tiwari DK, Reddy GV, Giri PV, Kumar N, Malla R, Rachel V, Bramhachari PV, et al. Anthocyanins: Multi-target agents for prevention and therapy of chronic diseases. Cur Pharm Des, 2017:**23**(41):6321-6346. https://doi.org/10.2174/1381612823666170519151801

Riehle MM, Bennett AF, Long AD. Genetic architecture of thermal adaptation in *Escherichia coli*. Proc Natl Acad Sci U S A, 2001:**98**(2):525-530. https://doi.org/10.1073/pnas.98.2.525

Rizzon C, Ponger L, Gaut BS. Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. PLoS Comput Biol, 2006:**2**(9):115. https://doi.org/10.1371/journal.pcbi.0020115

Saigo T, Wang T, Watanabe M, Tohge T. Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. Curr Opin Plant Biol, 2020:**55**:93-99. https://doi.org/10.1016/j.pbi.2020.04.001

Sandve SR, Rohlfs RV, Hvidsten TR. Subfunctionalization versus neofunctionalization after whole-genome duplication. Nat Genet, 2018:**50**(7):908-909. https://doi.org/10.1038/s41588-018-0162-4

Shin DH, Choi MG, Kang CS, Park CS, Choi SB, Park YI. A wheat R2R3-MYB protein *PURPLE PLANT1* (*TaPL1*) functions as a positive regulator of anthocyanin biosynthesis. Biochem Biophys Res Commun, 2016:**469**(3):686-691. https://doi.org/10.1016/j.bbrc.2015.12.001

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. Curr Opin Genet Dev, 2015:**35**:119-125. https://doi.org/10.1016/j.gde.2015.11.003

Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol, 2016:**30**:159-165. https://doi.org/10.1016/j.pbi.2016.03.015

Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in *Arabidopsis*

thaliana. Curr Opin Plant Biol, 2001:**4**(5):447-456. https://doi.org/10.1016/s1369-5266(00)00199-0

Sundaramoorthy J, Park GT, Lee JD, Kim JH, Seo HS, Song JT. Genetic and molecular regulation of flower pigmentation in soybean. J KOREAN SOC APPL BI, 2015: **58**(4):555-562. https://doi.org/10.1007/s13765-015-0077-z

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 1994:**22**(22):4673-4680. https://doi.org/10.1093/nar/22.22.4673

Wang CN, Ji WK, Liu YC, Zhou P, Meng YY, Zhang PC, Wen JQ, Mysore KS, Zhai JX, Young ND, et al. The antagonistic MYB paralogs *RH1* and *RH2* govern anthocyanin leaf markings in *Medicago truncatula*. New Phytol, 2021:**229**(6):3330-3344. https://doi.org/10.1111/nph.17097

Wang JP, Sun PC, Li YX, Liu YZ, Yu JG, Ma XL, Sun SR, Yang NS, Xia RY, Lei TY, et al. Hierarchically aligning 10 legume genomes establishes a family-level genomics platform. Plant Physiol, 2017:**174**(1):284-300. https://doi.org/10.1104/pp.16.01981.

Wang S, Xu Z, Yang Y, Ren W, Fang J, Wan L. Genome-wide analysis of R2R3-MYB genes in cultivated peanut (*Arachis hypogaea* L.): Gene duplications, functional conservation, and diversification. Front Plant Sci, 2023:**14**:1102174 https://doi.org/1102174. 10.3389/fpls.2023.1102174

Wang XP, Wang W, Chen SY, Lian YJ, Wang SC. Tropaeolum majus R2R3 MYB transcription factor *TmPAP2* functions as a positive regulator of anthocyanin biosynthesis. Int J Mol Sci, 2022:**23**(20):12395. https://doi.org/10.3390/ijms232012395

Xie M, Chung CY, Li MW, Wong FL, Wang X, Liu A, Wang Z, Leung AK, Wong TH, Tong SW, et al. A reference-grade wild soybean genome. Nat Commun, 2019:**10**(1):1216. https://doi.org/10.1038/s41467-019-09142-9

Xu WJ, Dubos C, Lepiniec, L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends in Plant Sci, 2015:**20**(3):176-185.

https://doi.org/10.1016/j.tplants.2014.12.001

Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol, 2007:**24**(8):1586-1591. https://doi.org/10.1093/molbev/msm088

Yin LW, Xu G, Yang JL, Zhao MX. The heterogeneity in the landscape of gene dominance in maize is accompanied by unique chromatin environments. Mol Biol Evol, 2022:**39**(10). https://doi.org/10.1093/molbev/msac198

Zabala G, Vodkin LO. Methylation affects transposition and splicing of a large CACTA transposon from a MYB transcription factor regulating anthocyanin synthase genes in soybean seed coats. PLoS One, 2014:**9**(11):111959. https://doi.org/10.1371/journal.pone.0111959

Zhang H, Zhang F, Yu YM, Feng L, Jia JB, Liu B, Li BS, Guo HW, Zhai JX. A comprehensive online database for exploring ~20,000 public *Arabidopsis* RNA-seq libraries. Mol Plant, 2020:**13**(9):1231-1233. https://doi.org/10.1016/j.molp.2020.08.001

Zhang N, McHale LK, Finer JJ. Isolation and characterization of "GmScream" promoters that regulate highly expressing soybean (*Glycine max.* Merr.) genes. Plant Sci, 2015:**241**:189-198. https://doi.org/10.1016/j.plantsci.2015.10.010

Zhang P, Chopra S, Peterson T. A segmental gene duplication generated differentially expressed myb-homologous genes in maize. Plant Cell, 2000:**12**(12):2311-2322. https://doi.org/ 10.1105/tpc.12.12.2311

Zhang SL, Kondorosi É, Kereszt A. An anthocyanin marker for direct visualization of plant transformation and its use to study nitrogen-fixing nodule development. J Plant Res, 2019:**132**(5):695-703. https://doi.org/10.1007/s10265-019-01126-6

Zhang Y, Butelli E, Martin C. Engineering anthocyanin biosynthesis in plants. Curr Opin Plant Biol, 2014:**19**:81-90. https://doi.org/10.1016/j.pbi.2014.05.011

Zhao MX, Zhang B, Lisch D, Ma JX. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. Plant Cell, 2017:**29**(12):2974-2994. https://doi.org/10.1105/tpc.17.00595

Figure Legends

Figure 1. Micro-synteny analysis of the *MYB* gene cluster in legumes.

(A) *GmMYBA5* (*Glyma.09g234900*), *GmMYBA2* (*Glyma.09g235100*), *GmMYBA1* (*Glyma.09g235300*) and the pseudogene (*Glyma.09g235000*) are located in a cluster on chromosome 9.

(B) Micro-synteny analysis of the *MYB* gene cluster in legumes. Each triangle represents a gene and the tip of the triangle indicates the direction of the gene. Genes belonging to the same family are depicted using the same color.

Figure 2. GmMYBA5, GmMYBA2 and GmMYBA1 are closely clustered and phylogenetically related in the soybean genome.

(A) An amino acid sequence alignment of GmMYBA5, GmMYBA2 and GmMYBA1. The conserved KPRPR[S/T] [F/L] motif in subgroup 6 involved in anthocyanin regulation (activation) is highlighted in a red box.

(B) Phylogenetic analysis of *MYB* genes involved in the anthocyanin and proanthocyanin pathways. The protein sequences were used to construct the phylogenetic tree using the neighbor-joining method in MEGA7, with 1000 bootstrap replicates. The numbers displayed at each node represent the bootstrap values that support the corresponding node, with vales above 50% from 1000 replicates being shown. The GenBank accession numbers corresponding to the MYB proteins are listed in Table S5.

Figure 3. *GmMYBA5*, *GmMYBA2* and *GmMYBA1* function as transcriptional activators.

(A) Subcellular localization of GmMYBA5, GmMYBA2 and GmMYBA1 in tobacco (*Nicotiana benthamiana*) leaf epidermal cells. Bars = 25 μm.

(B) Transactivation activity assays in yeast demonstrated strong transactivation activity for GmMYBA5, GmMYBA2 and GmMYBA1.

(C) Phenotype characteristics of *Arabidopsis* transgenic lines and wild type controls.

772 WT, wild type; A5, 35S:*GmMYBA5*; A2, 35S:*GmMYBA2*; A1, 35S:*GmMYBA1*. Bars
773 = 0.15 cm.
774
775 **Figure 4. Functional divergence of the duplicated *MYB* genes.**
776 (A) Gene expression analysis of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* in different
777 tissues of soybean.
778 (B) *Nicotiana benthamiana* leaves and total anthocyanin after the transient
779 overexpression of *GmMYBA5*, *GmMYBA2*, *GmMYBA1* and the control with an empty
780 vector (EV). Bar = 1 cm.
781 (C) Overexpression of *GmMYBA5*, *GmMYBA2* and *GmMYBA1* induced significant
782 enrichment of anthocyanins. The relative anthocyanin contents were quantified using
783 the formula (A530-0.25×A657)/fresh weight, representing one anthocyanin unit. The
784 data indicates the mean ± SD for three biological replicates. Statistical significance
785 was determined using Student's *t* test (***, $p<0.001$).
786
787 **Figure 5. Molecular basis underlying the functional divergence of the *MYB***
788 **genes.**
789 (A) Chimeric proteins generated by fusing the DNA binding domains (BD) and
790 activating domains (AD) from *GmMYBA5*, *GmMYBA2*, and *GmMYBA1*.
791 (B) *Nicotiana benthamiana* leaves and total anthocyanin after the transient expression
792 of different chimeric proteins. Bar = 1 cm.
793 (C) Transgenic soybean hairy roots of Wm82 overexpressing empty vector (EV) and
794 *GmMYBA5*, *GmMYBA2* and *GmMYBA1* with pM4. Bar = 0.25 cm.
795 (D) Expression analyses of anthocyanin-related genes in transgenic soybean hairy
796 roots of Wm82 overexpressing empty vector (EV) and *GmMYBA5*, *GmMYBA2* and
797 *GmMYBA1* with pM4. Different lowercase letters indicate significant differences
798 among groups based on Fisher's Least Significant Difference (LSD) test at $P < 0.05$.
799
800 **Figure 6. Metabolomics analysis of soybean hairy roots overexpressing the *MYB***
801 **genes revealed distinct types and abundances of downstream metabolites.**

(A) Classification and number of all flavonoids detected across the samples.

(B) Number of differential metabolites between all comparisons. The numbers of up-regulated metabolites and down-regulated metabolites are shown in the histogram. EV, empty vector; A5, pM4:*GmMYBA5*; A2, pM4:*GmMYBA2*; A1, pM4:*GmMYBA1*.

(C) Venn analysis depicting the overlapping and unique differential metabolites in hairy roots overexpressing *GmMYBA5, GmMYBA2* and *GmMYBA1*. A5, pM4:*GmMYBA5*; A2, pM4:*GmMYBA2*; A1, pM4:*GmMYBA1*.

(D) Types and changes of anthocyanins detected in the overexpression lines of *GmMYBA5, GmMYBA2*, and *GmMYBA1*. Glycoside derivatives are highlighted in red.

Figure 7. The model of the tandemly duplicated *MYB* genes functionally diverged in the regulation of anthocyanin biosynthesis pathway.

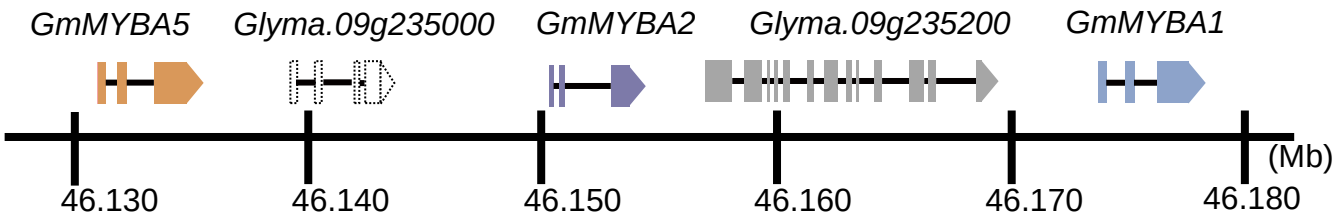
The *MYB* genes following tandem duplication undergone functional divergence, leading to tissue-specific expression patterns and differential activation abilities for anthocyanin biosynthesis pathway. CHI, chalcone isomerase; DFR, dihydroflavonol-4-reductase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3'5'hydroxylase; F3H, flavanone 3-hydroxylase; DFR, Dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; UFGT, UDP-flavonoid glucosyltransferase.

Table 1. Evolutionary distances of the three MYB genes within soybean and between soybean and common bean

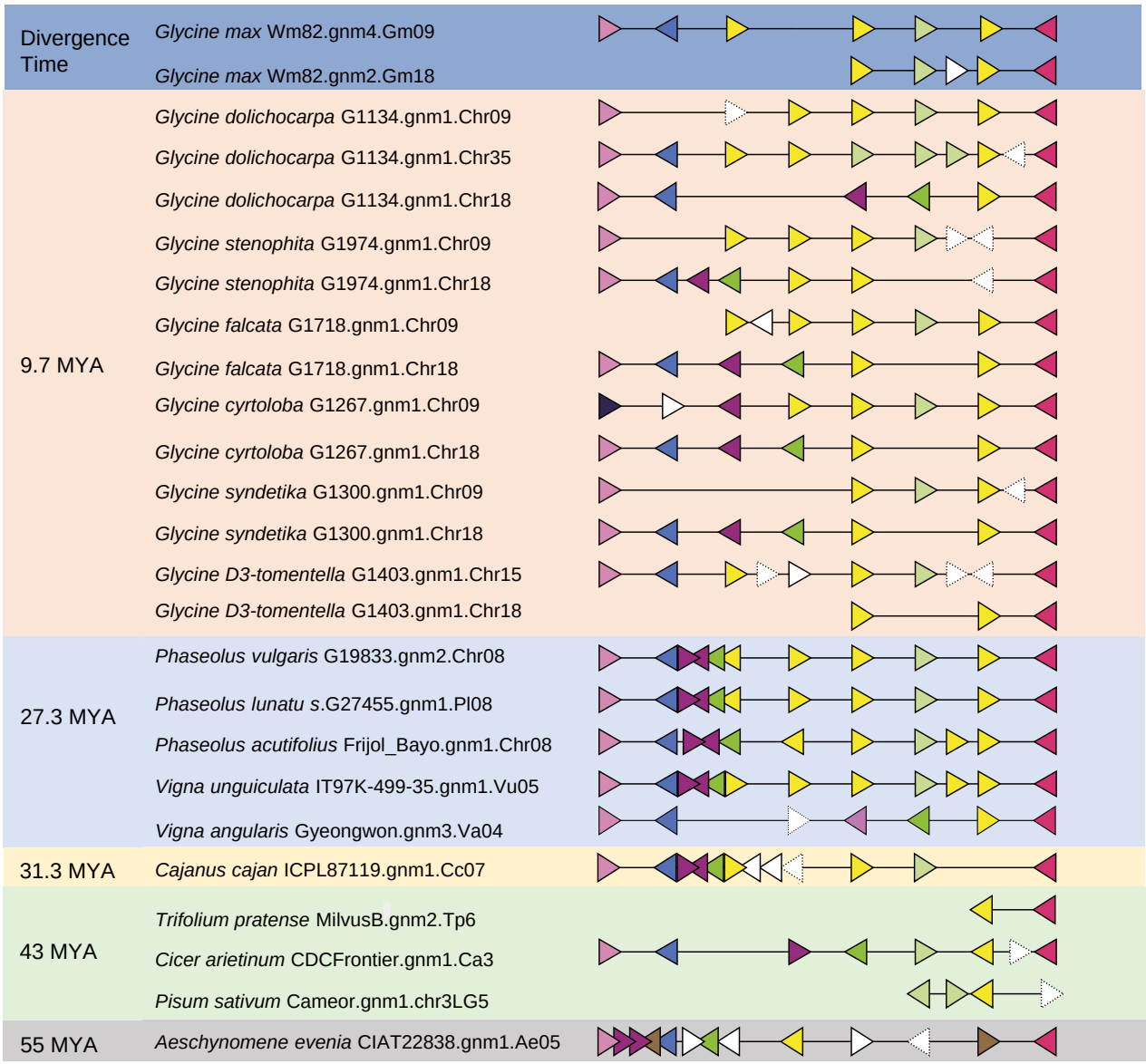
MYB gene	Homologous gene in soybean or common bean	Ka	Ks	ω (Ka/Ks)
Glyma.09G234900 (*GmMYBA5*)	*Phvul.008G038200*	0.1823	0.3714	0.4907
Glyma.09G235100 (GmMYBA2)	*Phvul.008G038200*	0.1776	0.3972	0.4471
Glyma.09G235300 (GmMYBA1)	*Phvul.008G038200*	0.1843	0.4659	0.3956
Glyma.09G234900 (*GmMYBA5*)	*Glyma.09G235100 (GmMYBA2)*	0.0758	0.2030	0.3735
Glyma.09G234900 (*GmMYBA5*)	*Glyma.09G235300 (GmMYBA1)*	0.0940	0.2758	0.3407
Glyma.09G235100 (GmMYBA2)	*Glyma.09G235300 (GmMYBA1)*	0.0880	0.2721	0.3234
Phvul.008G038200	*Phvul.008G038400*	0.1740	0.4100	0.4244
Phvul.008G038200	*Phvul.008G038500*	0.2020	0.6456	0.3128
Phvul.008G038200	*Phvul.008G038600*	0.0356	0.0654	0.5444
Phvul.008G038400	*Phvul.008G038500*	0.0821	0.1414	0.5810
Phvul.008G038400	*Phvul.008G038600*	0.1904	0.5139	0.3705
Phvul.008G038500	*Phvul.008G038600*	0.2131	0.6243	0.3413

Fig.1

A



B



- MYB transcription factor
- Phosphate transporter PHO1 homolog
- Pentatricopeptide repeat (PPR) superfamily protein
- Vesicle transport v-SNARE family protein
- Cystathionine Gamma-Synthase
- emp24/gp25L/p24 family/GOLD family protein
- Singletons
- Orphans
- Pectinesterase
- Protein of unknown function

Fig.2

A

B

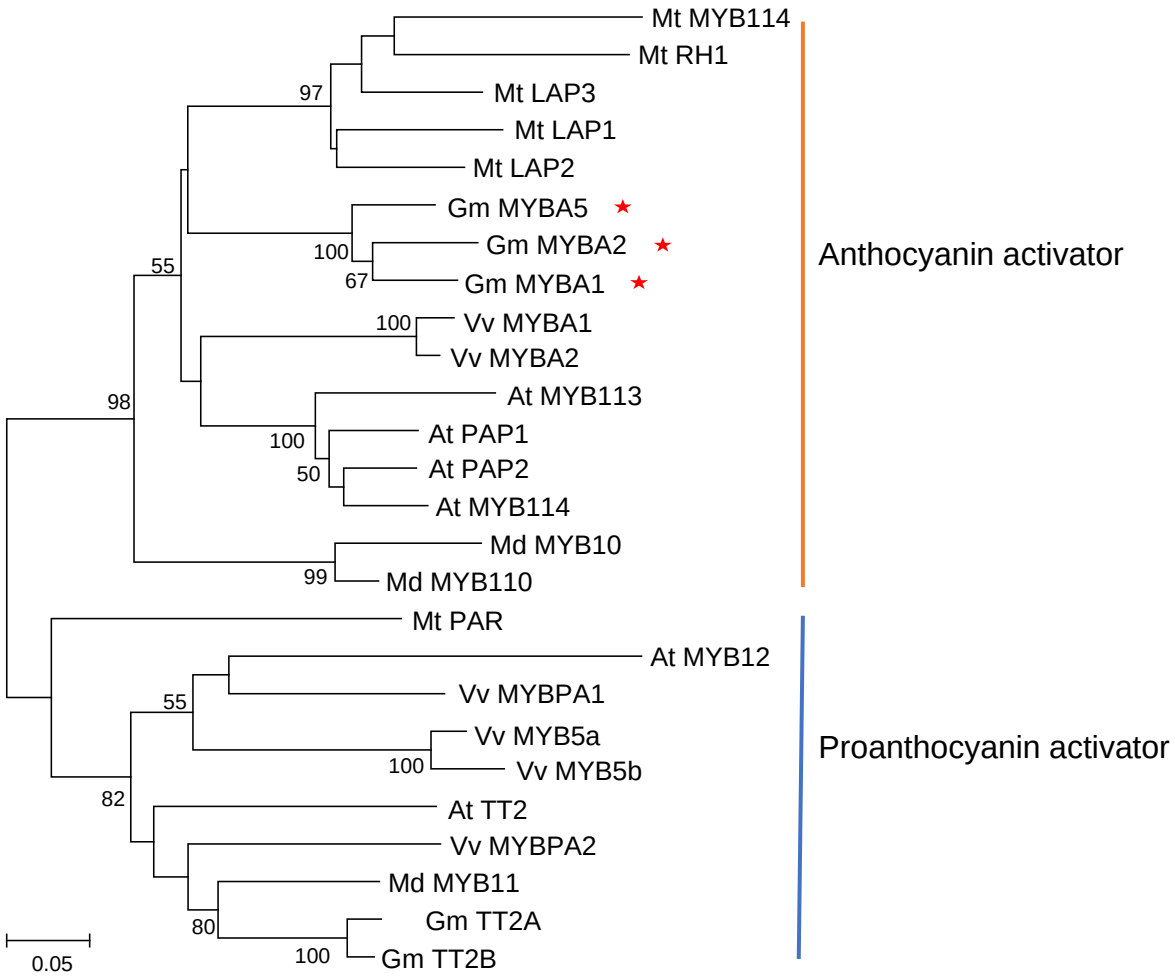


Fig.3

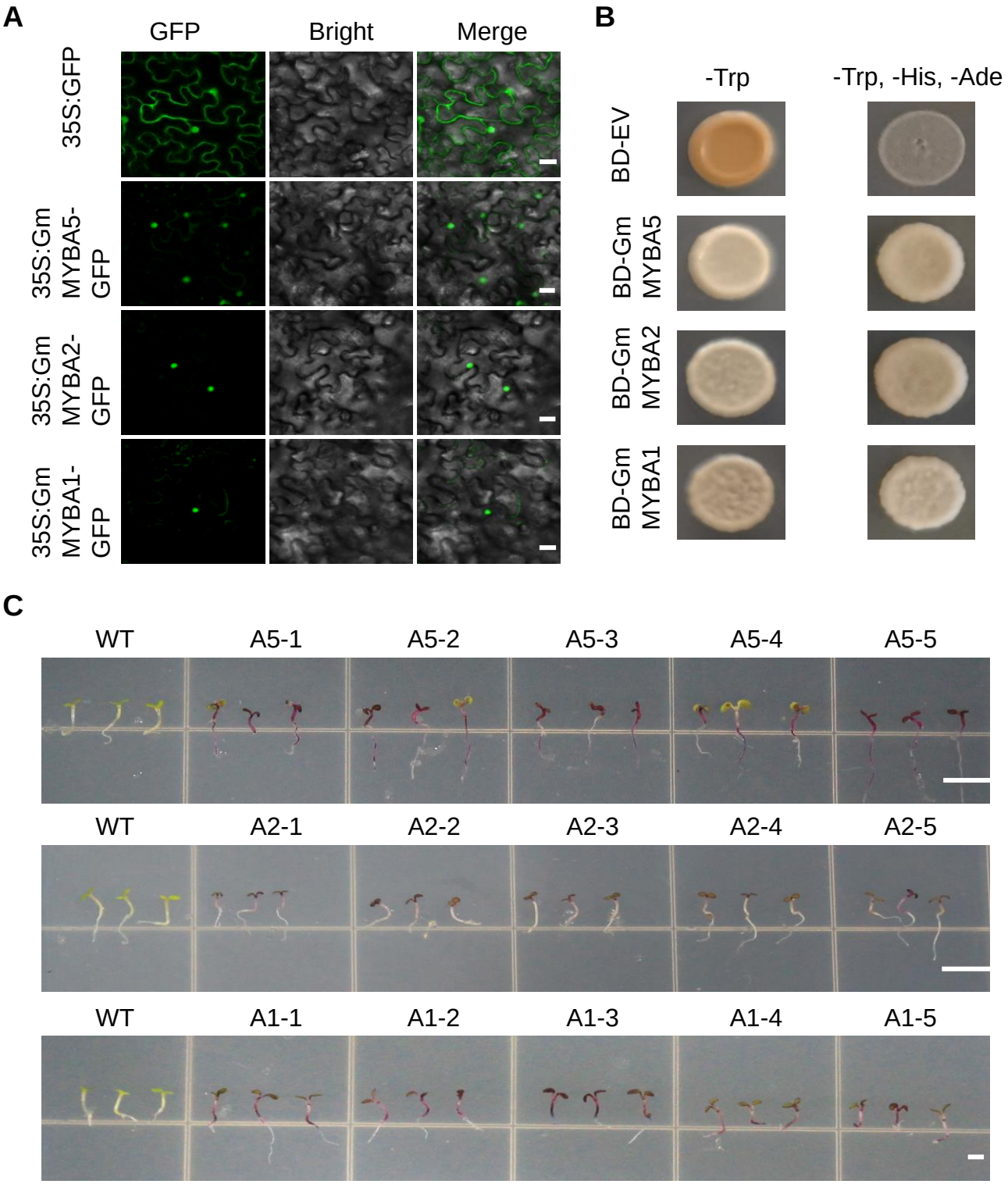


Fig.4

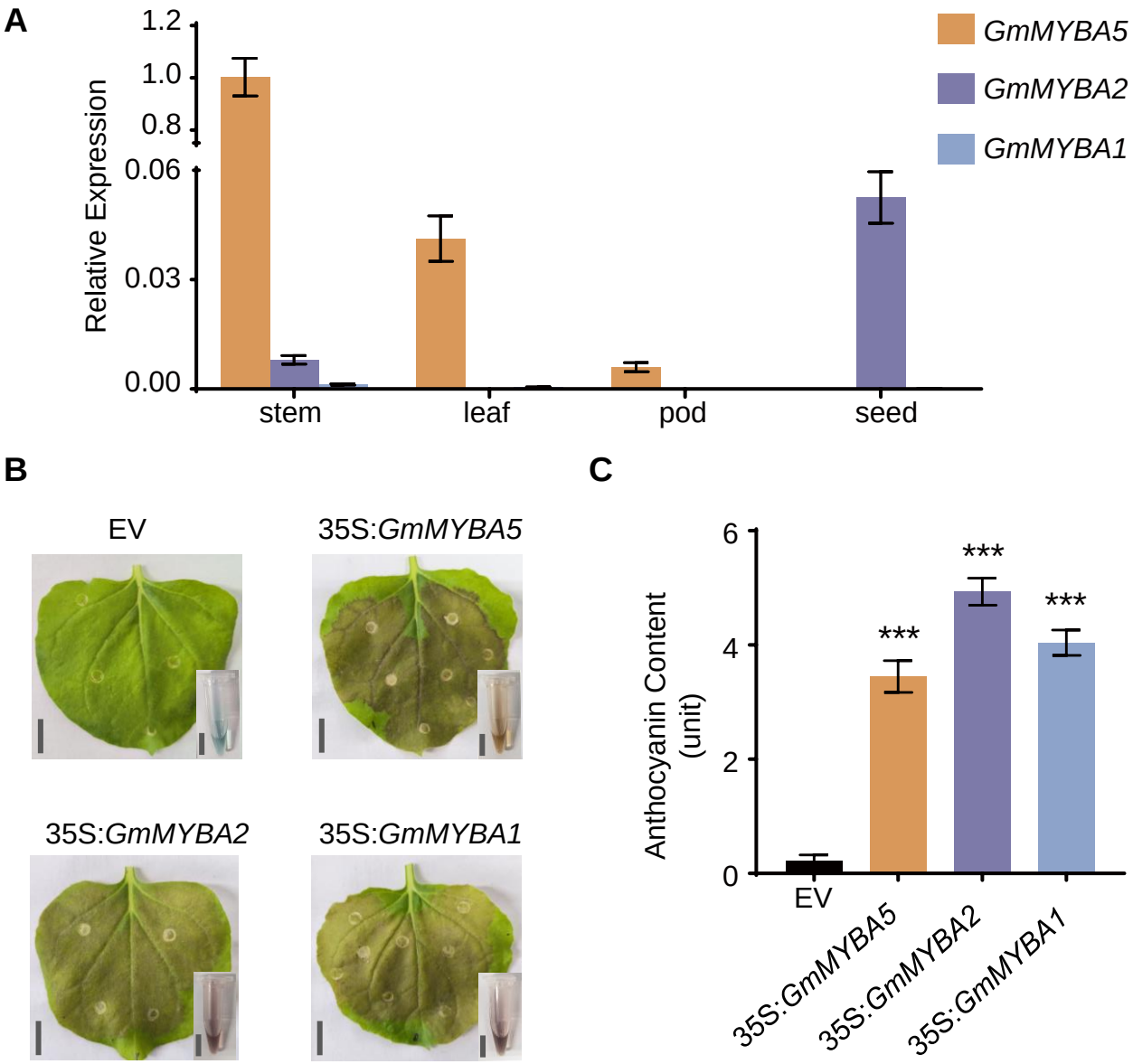


Fig.5

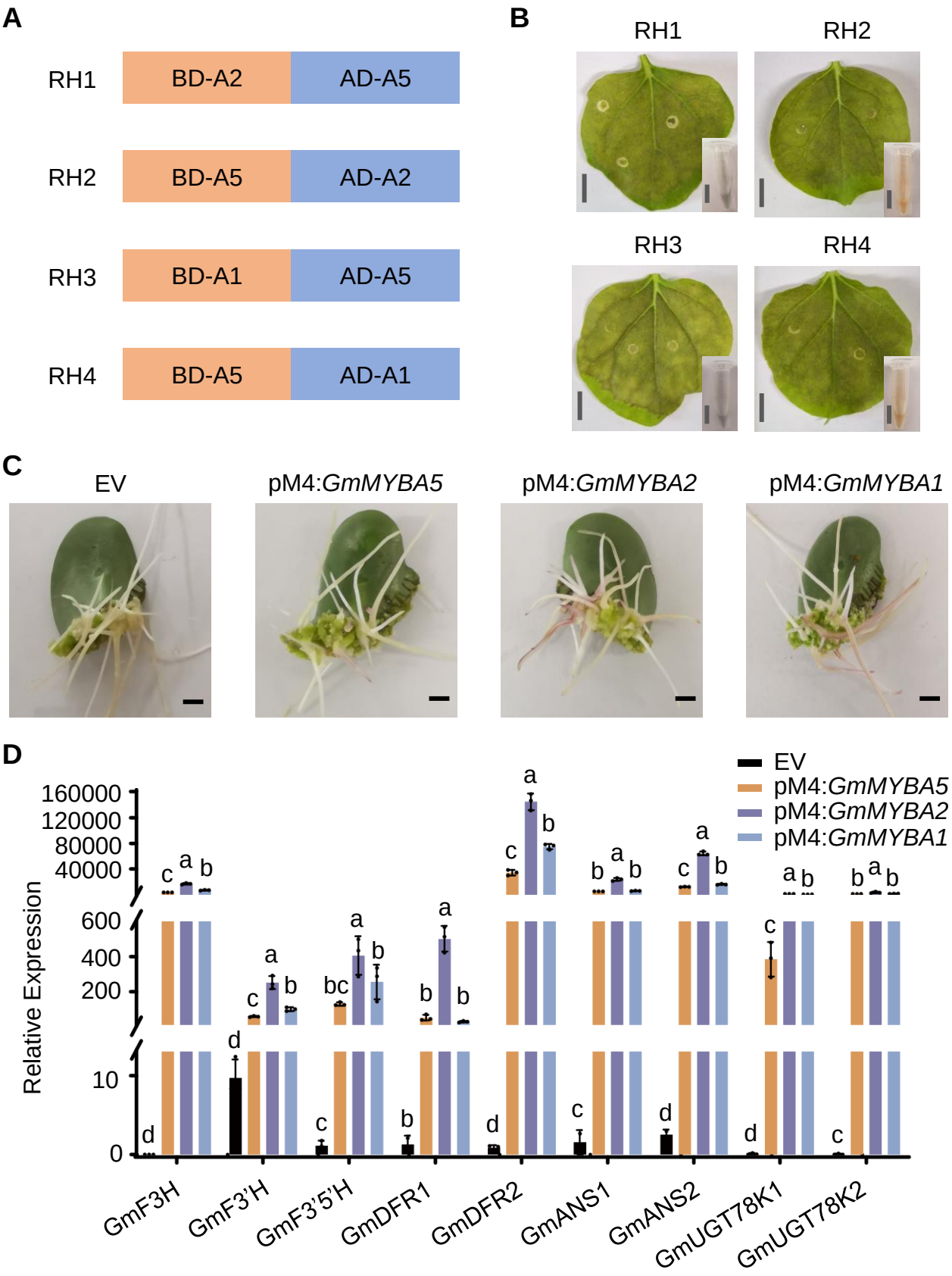


Fig.6

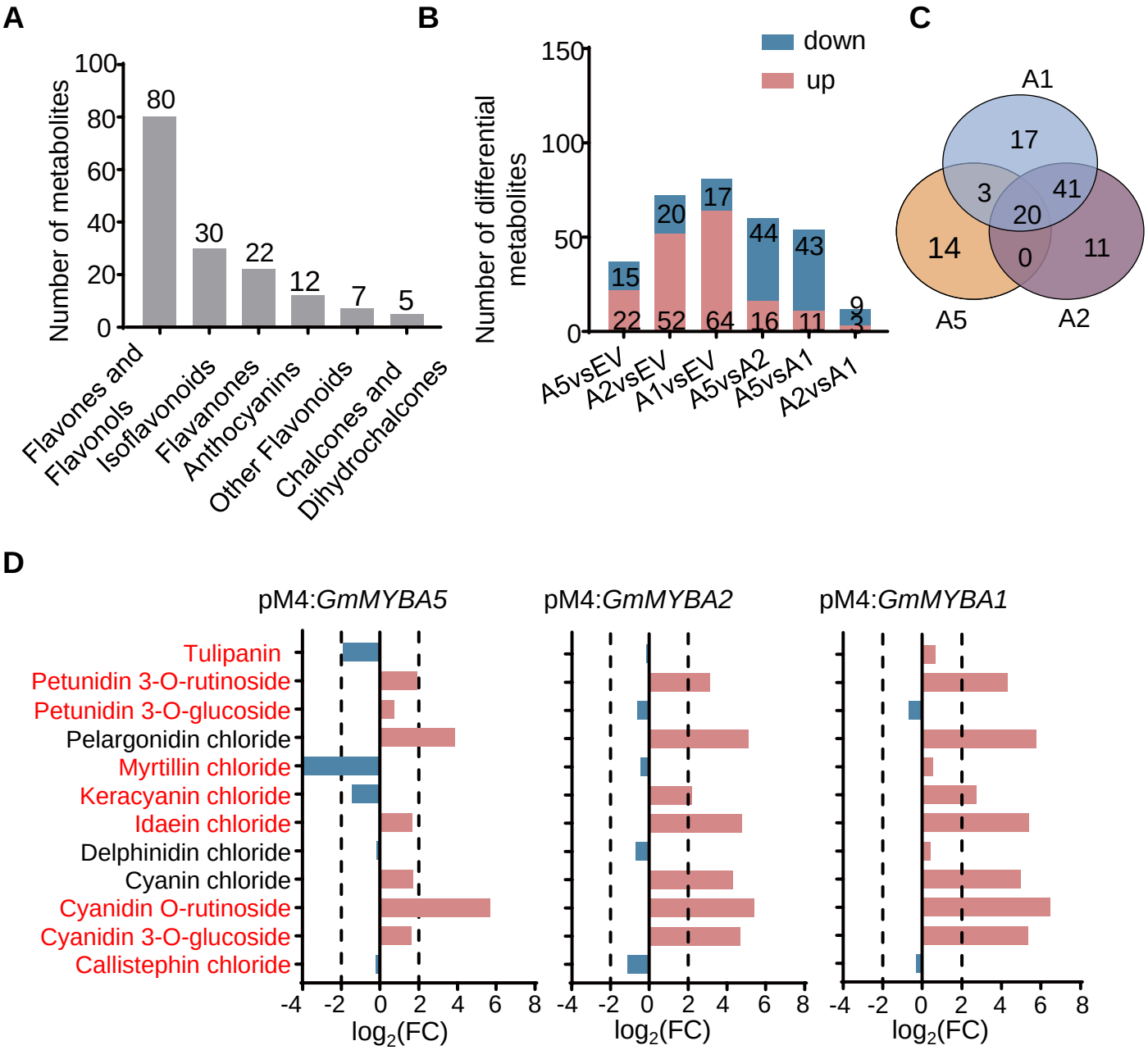
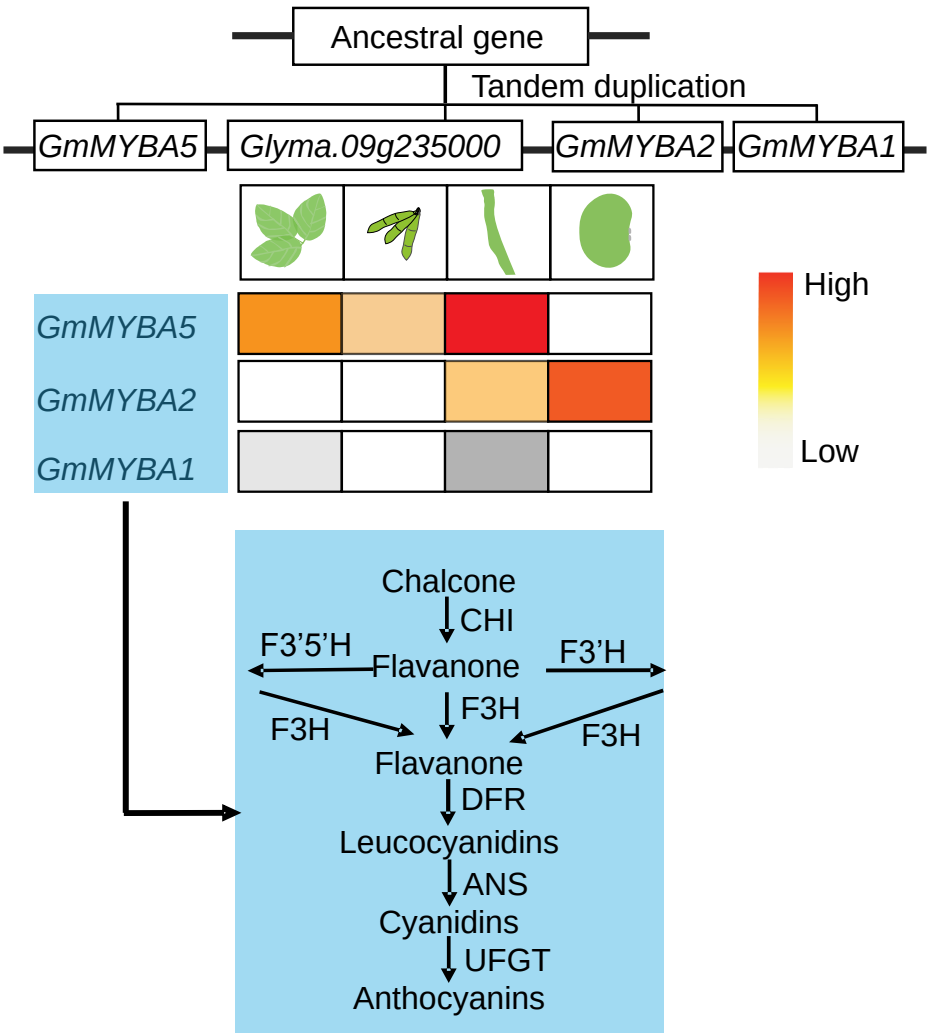


Fig.7



Parsed Citations

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell. 2014:26(3):962-980. https://doi.org/10.1105/tpc.113.122069
Google Scholar: Author Only Title Only Author and Title

Bent A. Arabidopsis thaliana floral dip transformation method. Methods Mol Biol, 2006:343:87-103. https://doi.org/10.1385/1-59745-130-4:87
Google Scholar: Author Only Title Only Author and Title

Birchler JA, Veitia RA. One hundred years of gene balance: how stoichiometric issues affect gene expression, genome evolution, and quantitative traits. Cytogenet Genome Res, 2021:161(10-11):529-550. https://doi.org/10.1159/000519592
Google Scholar: Author Only Title Only Author and Title

Brown CJ, Todd KM, Rosenzweig RF. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol, 1998:15(8):931-942. https://doi.org/10.1093/oxfordjournals.molbev.a026009
Google Scholar: Author Only Title Only Author and Title

Chen C, Liu XQ, Li SZ, Liu CX, Zhang YL, Luo LL, Miao LQ, Yang WZ, Xiao ZJ, Zhong Y, et al. Co-expression of transcription factors ZmC1 and ZmR2 establishes an efficient and accurate haploid embryo identification system in maize. Plant J, 2022a:111(5):1296-1307. https://doi.org/10.1111/tpj.15888
Google Scholar: Author Only Title Only Author and Title

Chen Y, Fang T, Su H, Duan SF, Ma RR, Wang P, Wu L, Sun WB, Hu QC, Zhao MX, et al. A reference-grade genome assembly for Astragalus mongholicus and insights into the biosynthesis and high accumulation of triterpenoids and flavonoids in its roots. Plant Commun, 2022b:4:100469. https://doi.org/10.1016/j.xplc.2022.100469
Google Scholar: Author Only Title Only Author and Title

Chin HS, Wu YP, Hour AL, Hong CY, Lin YR. Genetic and evolutionary analysis of purple leaf sheath in rice. Rice, 2016:9:8. https://doi.org/10.1186/s12284-016-0080-y
Google Scholar: Author Only Title Only Author and Title

D'Amelia V, Aversano R, Ruggiero A, Batelli G, Appelhagen I, Dinacci C, Hill L, Martin C, Carputo D. Subfunctionalization of duplicate MYB genes in Solanum commersonii generated the cold-induced ScAN2 and the anthocyanin regulator ScAN1. Plant Cell and Environ, 2018: 41(5):1038-1051. https://doi.org/10.1111/pce.12966
Google Scholar: Author Only Title Only Author and Title

Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev of Plant Biol, 2009:60:433-453. https://doi.org/10.1146/annurev.arplant.043008.092122
Google Scholar: Author Only Title Only Author and Title

Gao RF, Han TT, Xun HW, Zeng XS, Li PH, Li YQ, Wang YN, Shao Y, Cheng X, Feng XZ, et al. MYB transcription factors GmMYBA2 and GmMYBR function in a feedback loop to control pigmentation of seed coat in soybean. J Exp Bot, 2021:72(12):4401-4418. https://doi.org/10.1093/jxb/erab152
Google Scholar: Author Only Title Only Author and Title

García-Fernández C, Campa A, Ferreira JJ. Dissecting the genetic control of seed coat color in a RIL population of common bean (Phaseolus vulgaris L.). Theor and Appl Genet, 2021:134(11):3687-3698. https://doi.org/10.1007/s00122-021-03922-y
Google Scholar: Author Only Title Only Author and Title

Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc, 2007:2(1):31-34. https://doi.org/10.1038/nprot.2007.13
Google Scholar: Author Only Title Only Author and Title

Gillman JD, Tetlow A, Lee JD, Shannon JG, Bilyeu K. Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. BMC Plant Biol, 2011:11:155 https://doi.org/155. 10.1186/1471-2229-11-155
Google Scholar: Author Only Title Only Author and Title

Gould KS. Nature's Swiss Army Knife: The diverse protective roles of anthocyanins in leaves. J of Biomed and Biotechnol, 2004:2004(5):314-320. https://doi.org/10.1155/S1110724304406147
Google Scholar: Author Only Title Only Author and Title

Grotewold E, Drummond BJ, Bowen B, Peterson T. The MYB-homologous p-gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell, 1994:76(3):543-553. https://doi.org/ 10.1016/0092-8674(94)90117-1
Google Scholar: Author Only Title Only Author and Title

Guo WB, Zhao J, Li XX, Qin L, Yan XL, Liao H. A soybean β-expansin gene GmEXPB2 intrinsically involved in root system

architecture responses to abiotic stresses. Plant J, 2011:66(3):541-552. https://doi.org/10.1111/j.1365-313X.2011.04511.x
Google Scholar: Author Only Title Only Author and Title

Hastings PJ. Adaptive amplification. Crit Rev Biochem and Mol Biol, 2007:42(4):271-283. https://doi.org/10.1080/10409230701507757
Google Scholar: Author Only Title Only Author and Title

Herniter IA, Muñoz-Amatriaín M, Lo S, Guo YN, Close TJ. Identification of candidate genes controlling black seed coat and pod tip color in cowpea (Vigna unguiculata [L.] Walp). G3 (Bethesda), 2018:8(10):3347-3355. https://doi.org/10.1534/g3.118.200521
Google Scholar: Author Only Title Only Author and Title

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J Exp Bot, 2011:62(8):2465-2483. https://doi.org/10.1093/jxb/erq442
Google Scholar: Author Only Title Only Author and Title

Hou JJ, Fan WW, Ma RR, Li B, Yuan ZH, Huang WX, Wu YY, Hu Q, Lin CJ, Zhao XQ, et al. MALE STERILITY 3 encodes a plant homeodomain-finger protein for male fertility in soybean. J Integr Plant Biol, 2022:64(5):1076-1086. https://doi.org/10.1111/jipb.13242
Google Scholar: Author Only Title Only Author and Title

Huang D, Wang X, Tang ZZ, Yuan Y, Xu YT, He JX, Jiang XL, Peng SA, Li L, Butelli E, et al. Subfunctionalization of the Ruby2-Ruby1 gene cluster during the domestication of citrus. Nat Plants, 2018:4(11):930-941. https://doi.org/10.1038/s41477-018-0287-6
Google Scholar: Author Only Title Only Author and Title

Huang T, Xin S, Fang Y, Chen T, Chang J, Ko N. Huang H, Hua Y. Use of a novel R2R3-MYB transcriptional activator of anthocyanin biosynthesis as visual selection marker for rubber tree (Hevea brasiliensis) transformation. Ind Crop and Prod, 2021:174:114225. https://doi.org/10.1016/j.indcrop.2021.114225
Google Scholar: Author Only Title Only Author and Title

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NW, Couloux A, Dalwani A, Denny R, et al. Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiology, 2008:148(4):1740-1759. https://doi.org/10.1104/pp.108.127902
Google Scholar: Author Only Title Only Author and Title

James TC, Usher J, Campbell S, Bond U. Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. Curr Genet, 2008:53(3):139-152. https://doi.org/10.1007/s00294-007-0172-8
Google Scholar: Author Only Title Only Author and Title

Jeong SC, Moon JK, Park SK, Kim MS, Lee K, Lee SR, Jeong N, Choi MS, Kim N, Kang ST, et al. Genetic diversity patterns and domestication origin of soybean. Theor Appl Genet, 2019:132(4):1179-1193. https://doi.org/10.1007/s00122-018-3271-7
Google Scholar: Author Only Title Only Author and Title

Kereszt A, Li D, Indrasumunar A, Nguyen CD, Nontachaiyapoom S, Kinkema M, Gresshoff PM. Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. Nat Protoc, 2007:2(4):948-952. https://doi.org/10.1038/nprot.2007.141
Google Scholar: Author Only Title Only Author and Title

Kim MY, Van K, Kang YJ, Kim KH, Lee SH. Tracing soybean domestication history: From nucleotide to genome. Breeding Sci, 2012:61(5):445-452. https://doi.org/10.1270/jsbbs.61.445
Google Scholar: Author Only Title Only Author and Title

Kovinich N, Kayanja G, Chanoca A, Otegui MS, Grotewold E. Abiotic stresses induce different localizations of anthocyanins in Arabidopsis. Plant Signal Behav, 2015:10(7):1027850. https://doi.org/10.1080/15592324.2015.1027850
Google Scholar: Author Only Title Only Author and Title

Li CX, Yu WJ, Xu JR, Lu XF, Liu YZ. Anthocyanin biosynthesis induced by MYB transcription factors in plants. Int J Mol Sci, 2022a:23(19):11701. https://doi.org/10.3390/ijms231911701
Google Scholar: Author Only Title Only Author and Title

Li Y, Fang XJ, Lin ZW. Convergent loss of anthocyanin pigments is controlled by the same MYB gene in cereals. J Exp Bot, 2022b:73(18):6089-6102. https://doi.org/10.1093/jxb/erac270

Li YH, Guan RX, Liu ZX, Ma YS, Wang LX, Li LH, Lin FY, Luan WJ, Chen PY, Yan Z, et al. Genetic structure and diversity of cultivated soybean (Glycine max (L.) Merr.) landraces in China. Theor Appl Genet, 2008:117(6):857-871. https://doi.org/10.1007/s00122-008-0825-0
Google Scholar: Author Only Title Only Author and Title

Lim SH, Kim DH, Cho MC, Lee JY. Chili Pepper AN2 (CaAN2): A visible selection marker for nondestructive monitoring of transgenic plants. Plants (Basel), 2022.11(6):820. https://doi.org/10.3390/plants11060820
Google Scholar: Author Only Title Only Author and Title

Liu JY, Osbourn A, Ma PD. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. Mol Plant,

2015:8(5):689-708. https://doi.org/10.1016/j.molp.2015.03.012
Google Scholar: Author Only Title Only Author and Title

Lloyd A, Brockman A, Aguirre L, Campbell A, Bean A, Cantero A, Gonzalez A. Advances in the MYB-bHLH-WD Repeat (MBW) pigment regulatory model: Addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. Plant and Cell Physiol, 2017: 58(9):1431-1441. https://doi.org/10.1093/pcp/pcx075
Google Scholar: Author Only Title Only Author and Title

Lu N, Rao X, Li Y, Jun JH, Dixon RA. Dissecting the transcriptional regulation of proanthocyanidin and anthocyanin biosynthesis in soybean (Glycine max). Plant Biotechnol J, 2021:19(7):1429-1442. https://doi.org/10.1111/pbi.13562
Google Scholar: Author Only Title Only Author and Title

Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. J Genet, 2013:92(1):155-161. https://doi.org/10.1007/s12041-013-0212-8
Google Scholar: Author Only Title Only Author and Title

Malle S, Morrison M, Belzile F. Identification of loci controlling mineral element concentration in soybean seeds. BMC Plant Biol, 2020:20(1):419. https://doi.org/10.1186/s12870-020-02631-w
Google Scholar: Author Only Title Only Author and Title

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell, 2003:15(4):809-834. https://doi.org/10.1105/tpc.009308
Google Scholar: Author Only Title Only Author and Title

O'Neil J, Tchinda J, Gutierrez A, Moreau L, Maser RS, Wong KK, Li W, McKenna K, Liu XS, Feng B, et al. Alu elements mediate MYB gene tandem duplication in human T-ALL. J Exp Med, 2007:204(13):3059-3066. https://doi.org/10.1084/jem.20071637
Google Scholar: Author Only Title Only Author and Title

Putta S, Yarla NS, Peluso I, Tiwari DK, Reddy GV, Giri PV, Kumar N, Malla R, Rachel V, Bramhachari PV, et al. Anthocyanins: Multi-target agents for prevention and therapy of chronic diseases. Cur Pharm Des, 2017:23(41):6321-6346. https://doi.org/10.2174/1381612823666170519151801
Google Scholar: Author Only Title Only Author and Title

Riehle MM, Bennett AF, Long AD. Genetic architecture of thermal adaptation in Escherichia coli. Proc Natl Acad Sci U S A, 2001:98(2):525-530. https://doi.org/10.1073/pnas.98.2.525
Google Scholar: Author Only Title Only Author and Title

Rizzon C, Ponger L, Gaut BS. Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLoS Comput Biol, 2006:2(9):115. https://doi.org/10.1371/journal.pcbi.0020115
Google Scholar: Author Only Title Only Author and Title

Saigo T, Wang T, Watanabe M, Tohge T. Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. Curr Opin Plant Biol, 2020:55:93-99. https://doi.org/10.1016/j.pbi.2020.04.001
Google Scholar: Author Only Title Only Author and Title

Sandve SR, Rohlfs RV, Hvidsten TR. Subfunctionalization versus neofunctionalization after whole-genome duplication. Nat Genet, 2018:50(7):908-909. https://doi.org/10.1038/s41588-018-0162-4
Google Scholar: Author Only Title Only Author and Title

Shin DH, Choi MG, Kang CS, Park CS, Choi SB, Park YI. A wheat R2R3-MYB protein PURPLE PLANT1 (TaPL1) functions as a positive regulator of anthocyanin biosynthesis. Biochem Biophys Res Commun, 2016:469(3):686-691. https://doi.org/10.1016/j.bbrc.2015.12.001
Google Scholar: Author Only Title Only Author and Title

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. Curr Opin Genet Dev, 2015:35:119-125. https://doi.org/10.1016/j.gde.2015.11.003
Google Scholar: Author Only Title Only Author and Title

Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol, 2016:30:159-165. https://doi.org/10.1016/j.pbi.2016.03.015
Google Scholar: Author Only Title Only Author and Title

Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol, 2001:4(5):447-456. https://doi.org/10.1016/s1369-5266(00)00199-0
Google Scholar: Author Only Title Only Author and Title

Sundaramoorthy J, Park GT, Lee JD, Kim JH, Seo HS, Song JT. Genetic and molecular regulation of flower pigmentation in soybean. J KOREAN SOC APPL BI, 2015: 58(4):555-562. https://doi.org/10.1007/s13765-015-0077-z
Google Scholar: Author Only Title Only Author and Title

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 1994:22(22):4673-4680. https://doi.org/10.1093/nar/22.22.4673
Google Scholar: Author Only Title Only Author and Title

Wang CN, Ji WK, Liu YC, Zhou P, Meng YY, Zhang PC, Wen JQ, Mysore KS, Zhai JX, Young ND, et al. The antagonistic MYB paralogs RH1 and RH2 govern anthocyanin leaf markings in Medicago truncatula. New Phytol, 2021:229(6):3330-3344. https://doi.org/10.1111/nph.17097
Google Scholar: Author Only Title Only Author and Title

Wang JP, Sun PC, Li YX, Liu YZ, Yu JG, Ma XL, Sun SR, Yang NS, Xia RY, Lei TY, et al. Hierarchically aligning 10 legume genomes establishes a family-level genomics platform. Plant Physiol, 2017:174(1):284-300. https://doi.org/10.1104/pp.16.01981.
Google Scholar: Author Only Title Only Author and Title

Wang S, Xu Z, Yang Y, Ren W, Fang J, Wan L. Genome-wide analysis of R2R3-MYB genes in cultivated peanut (Arachis hypogaea L.): Gene duplications, functional conservation, and diversification. Front Plant Sci, 2023:14:1102174 https://doi.org/1102174. 10.3389/fpls.2023.1102174
Google Scholar: Author Only Title Only Author and Title

Wang XP, Wang W, Chen SY, Lian YJ, Wang SC. Tropaeolum majus R2R3 MYB transcription factor TmPAP2 functions as a positive regulator of anthocyanin biosynthesis. Int J Mol Sci, 2022:23(20):12395. https://doi.org/10.3390/ijms232012395
Google Scholar: Author Only Title Only Author and Title

Xie M, Chung CY, Li MW, Wong FL, Wang X, Liu A, Wang Z, Leung AK, Wong TH, Tong SW, et al. A reference-grade wild soybean genome. Nat Commun, 2019:10(1):1216. https://doi.org/10.1038/s41467-019-09142-9
Google Scholar: Author Only Title Only Author and Title

Xu WJ, Dubos C, Lepiniec, L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends in Plant Sci, 2015:20(3):176-185. https://doi.org/10.1016/j.tplants.2014.12.001
Google Scholar: Author Only Title Only Author and Title

Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol, 2007:24(8):1586-1591. https://doi.org/10.1093/molbev/msm088
Google Scholar: Author Only Title Only Author and Title

Yin LW, Xu G, Yang JL, Zhao MX. The heterogeneity in the landscape of gene dominance in maize is accompanied by unique chromatin environments. Mol Biol Evol, 2022:39(10). https://doi.org/10.1093/molbev/msac198
Google Scholar: Author Only Title Only Author and Title

Zabala G, Vodkin LO. Methylation affects transposition and splicing of a large CACTA transposon from a MYB transcription factor regulating anthocyanin synthase genes in soybean seed coats. PLoS One, 2014:9(11):111959. https://doi.org/10.1371/journal.pone.0111959
Google Scholar: Author Only Title Only Author and Title

Zhang H, Zhang F, Yu YM, Feng L, Jia JB, Liu B, Li BS, Guo HW, Zhai JX. A comprehensive online database for exploring ~20,000 public Arabidopsis RNA-seq libraries. Mol Plant, 2020:13(9):1231-1233. https://doi.org/10.1016/j.molp.2020.08.001
Google Scholar: Author Only Title Only Author and Title

Zhang N, McHale LK, Finer JJ. Isolation and characterization of "GmScream" promoters that regulate highly expressing soybean (Glycine max. Merr.) genes. Plant Sci, 2015:241:189-198. https://doi.org/10.1016/j.plantsci.2015.10.010
Google Scholar: Author Only Title Only Author and Title

Zhang P, Chopra S, Peterson T. A segmental gene duplication generated differentially expressed myb-homologous genes in maize. Plant Cell, 2000:12(12):2311-2322. https://doi.org/ 10.1105/tpc.12.12.2311
Google Scholar: Author Only Title Only Author and Title

Zhang SL, Kondorosi É, Kereszt A. An anthocyanin marker for direct visualization of plant transformation and its use to study nitrogen-fixing nodule development. J Plant Res, 2019:132(5):695-703. https://doi.org/10.1007/s10265-019-01126-6
Google Scholar: Author Only Title Only Author and Title

Zhang Y, Butelli E, Martin C. Engineering anthocyanin biosynthesis in plants. Curr Opin Plant Biol, 2014:19:81-90. https://doi.org/10.1016/j.pbi.2014.05.011
Google Scholar: Author Only Title Only Author and Title

Zhao MX, Zhang B, Lisch D, Ma JX. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. Plant Cell, 2017:29(12):2974-2994. https://doi.org/10.1105/tpc.17.00595
Google Scholar: Author Only Title Only Author and Title